

8-21-2018

## Methods for Estimating Vehicle Fuel Consumption in Urban Road Networks Using Geographical Attributes

Ryan G. Conrad

*Portland State University*, [rconrad514@gmail.com](mailto:rconrad514@gmail.com)

Let us know how access to this document benefits you.

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/cengin\\_gradprojects](https://pdxscholar.library.pdx.edu/cengin_gradprojects)



Part of the [Civil and Environmental Engineering Commons](#)

---

### Recommended Citation

Conrad, Ryan G., "Methods for Estimating Vehicle Fuel Consumption in Urban Road Networks Using Geographical Attributes" (2018). *Civil and Environmental Engineering Master's Project Reports*. 43.  
[https://pdxscholar.library.pdx.edu/cengin\\_gradprojects/43](https://pdxscholar.library.pdx.edu/cengin_gradprojects/43)

This Project is brought to you for free and open access. It has been accepted for inclusion in Civil and Environmental Engineering Master's Project Reports by an authorized administrator of PDXScholar. For more information, please contact [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

Methods for Estimating Vehicle Fuel Consumption in Urban Road Networks Using  
Geographical Attributes

BY

RYAN G. CONRAD

A research project submitted in partial fulfillment  
of the requirement for the degree of

MASTER OF SCIENCE

IN

CIVIL AND ENVIRONMENTAL ENGINEERING

Project Advisor:

Dr. Miguel Andres Figliozzi

Portland State University

©2018

## **ACKNOWLEDGMENTS**

It's been a long journey, but I finally made it. My sincere thanks and gratitude to Dr. Miguel Figliozi for his patience and guidance throughout graduate school and with helping me complete this paper. I owe much of what I've learned about conducting research and writing technical papers to him. Thanks also to Dr. Chris Monsere for his help, guidance and encouragement throughout the program. I would also like to recognize and thank the Office of Graduate Studies for allowing me to complete my degree on a slightly protracted schedule.

A huge amount of recognition and thanks is owed to my amazing and beautiful wife, who encouraged me to finish my degree, endured the late nights and long hours I spent at the library, and painstakingly grammar-checked this paper. I could not have done this without your help! You were the rallying cry to family and friends who continued to encourage me to finish, checked in with me as the work progressed, and took time out of their day to attend my presentation. My sincere thanks to all of you, because sometimes it really does take a village!

## ABSTRACT

This paper presents statistical techniques for estimating vehicle fuel consumption in urban road networks based on vehicle and geographical factors. A routing algorithm utilizing mapping data from OpenStreetMap and elevation data from the Shuttle Radar Topography Mission is presented and used to generate paths that minimize vehicle fuel consumption. The concept of a *fuel consumption estimating function* is proposed as an extension of the well-known distance-estimating functions that are widely used in logistics planning and research.

Statistical models are developed that estimate fuel consumption in three tested urban areas with vehicle weight, elevation and regional travel speed characteristics being the independent variables. The models were tested on measures drawn from the underlying graph data used by the pathfinding engine as well as those derived from measurements on a digital elevation model using common geographical information system tools. The results provide promising techniques for the estimation of vehicle fuel consumption using only geographical data for long-range planning purposes.

## TABLE OF CONTENTS

<b>1.0</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	BACKGROUND .....	1
1.2	RESEARCH CONTRIBUTION .....	1
<b>2.0</b>	<b>LITERATURE REVIEW .....</b>	<b>3</b>
2.1	CIRCUITY FACTORS AND DISTANCE ESTIMATING FUNCTIONS .....	3
2.2	FUEL-EFFICIENT ROUTING AND MODELING .....	4
<b>3.0</b>	<b>METHODOLOGY .....</b>	<b>6</b>
3.1	DATA SOURCES .....	6
3.1.1	<i>OpenStreetMap .....</i>	<i>6</i>
3.1.2	<i>Shuttle Radar Topography Mission .....</i>	<i>8</i>
3.2	SHORTEST PATH PROBLEM STATEMENT .....	9
3.2.1	<i>Dijkstra's Algorithm .....</i>	<i>10</i>
3.2.2	<i>A* Algorithm.....</i>	<i>10</i>
3.2.3	<i>Bi-directional algorithms.....</i>	<i>11</i>
3.2.4	<i>A* performance enhancements .....</i>	<i>12</i>
3.3	PATHFINDING ENGINE OVERVIEW .....	13
<b>4.0</b>	<b>ANALYSIS .....</b>	<b>15</b>
4.1	EXPERIMENTAL SETTING .....	15
4.1.1	<i>Fuel consumption calculation.....</i>	<i>15</i>

4.1.2	<i>Characteristics of Fuel-efficient Paths</i> .....	17
4.2	FUEL CONSUMPTION ESTIMATION MODELS .....	17
4.2.1	<i>Initial Development</i> .....	18
4.2.2	<i>Graph dataset models</i> .....	22
4.2.3	<i>SRTM dataset models</i> .....	24
4.3	EXPERIMENTAL RESULTS .....	27
4.3.1	<i>Benchmarking</i> .....	28
4.3.2	<i>Graph dataset results</i> .....	29
4.3.3	<i>SRTM dataset results</i> .....	30
<b>5.0</b>	<b>CONCLUSIONS</b> .....	<b>32</b>
5.1	CHALLENGES .....	32
5.2	RECOMMENDATIONS AND FUTURE WORK .....	32
<b>6.0</b>	<b>REFERENCES</b> .....	<b>34</b>
<b>7.0</b>	<b>APPENDIX</b> .....	<b>37</b>
7.1	GLOSSARY OF TERMS .....	37
7.2	URBAN TEST AREAS .....	38

## LIST OF TABLES

Table 1: OSM data elements and usage.....	8
Table 2: Comparison of hierarchical optimizations (HBA*) vs. regular bi-directional A* .....	13
Table 3: Vehicle fuel consumption parameters and values .....	16
Table 4: Travel speeds calculated by OpenStreetMap category (highway tag).....	17
Table 5: Comparison of key metrics for the three path optimization strategies .....	17
Table 6: Regression estimators and units.....	22
Table 7: Graph-based regression models.....	22
Table 8: Models tested on SRTM spatial data .....	27
Table 9: Urban area statistics.....	27
Table 10: Independent variable summary statistics .....	28
Table 11: Regression results for DEFs .....	29
Table 12: Regression results for FCEF derived from the graph dataset .....	30
Table 13: Regression results from SRTM dataset .....	31



## LIST OF FIGURES

Figure 1: XML for OpenStreetMap way .....	6
Figure 2: XML for a block of OpenStreetMap nodes .....	7
Figure 3: XML for OpenStreetMap relation (type restriction) .....	7
Figure 4: Key graph elements .....	7
Figure 5: SRTM file format. ....	9
Figure 6: Shortest path calculations with bi-directional Dijkstra's algorithm (left) and A* (right) with a Euclidean distance heuristic. Green dots indicate vertices explored by the forward search; red dots indicate those explored by the backwards search.....	11
Figure 7: Pathfinding engine data processing .....	14
Figure 8: Fuel consumed as a function of travel speed at different road grades .....	16
Figure 9: Fuel consumption vs. actual travel distance .....	18
Figure 10: Fuel consumption vs. Euclidean distance.....	19
Figure 11: Euclidean vs. Actual Travel Distance for fuel-efficient paths .....	19
Figure 12: Bounding box for a route with queried edges used to calculate $v_{st}$ and $\alpha_{st}$ . ....	21
Figure 13: Predicted vs. residual plots for Models 1-4 .....	24
Figure 14: Spatial view from SQL Server Management Studio showing line string object and intersecting SRTM coordinates. ....	25
Figure 15: Example query showing how routing and SRTM elevation are joined using SQL Server spatial functions.....	26
Figure 16: Portland, OR test area with path origin-destination locations.....	38

Figure 17: Eugene, OR test area. ....	39
Figure 18: Salem, OR test area. ....	40

## **1.0 INTRODUCTION**

### **1.1 Background**

Fuel-efficient pathfinding is a variant of the classic shortest path problem with widespread real-world applications. Significant research has been directed towards improving the search algorithms, modeling fuel consumption and examining the relationship between fuel usage and environmental variables such as roadway elevation profile and traffic congestion. Because vehicle emissions are often directly proportional to fuel consumption, research on minimizing fuel consumption can provide cost-saving benefits to commercial operators as well as help public agencies seeking to draft policies aimed at reducing air pollution.

The growth of research into fuel-efficient routing has also been matched by rapid growth of crowd-sourced mapping sources such as OpenStreetMap (OSM) that provide up-to-date and readily available map data. Elevation data has been available from the United States Geological Survey (USGS) for several years covering the U.S. More recently, the Shuttle Radar Topography Mission (SRTM) datasets have been made publicly available by NASA via the USGS and provide elevation data for most of the world.

### **1.2 Research Contribution**

Existing research shows that the elevation changes, travel time characteristics of the road network and vehicle weight have the largest impact on fuel consumption. Research investigating the statistical modeling of fuel consumption has focused mainly on the relationship between vehicle variables and consumption within the context of environmental conditions such as travel time and elevation change. Thus far, no research has attempted to generalize the impact of environmental variables, namely travel speed and elevation change, within the statistical models.

The primary contribution of this paper is to introduce techniques for estimating vehicle fuel consumption using topographical characteristics of the road network and the urban area in general. Intuitive statistical models are developed as an extension of commonly used distance-estimating functions (DEFs) that see widespread use in logistical applications. The derived Fuel Consumption Estimating Functions (FCEF) are tested on fuel-efficient paths calculated in three urban areas with

different topographical and travel speed characteristics. Tests are also conducted on two different datasets: one derived from the road network data used by the pathfinding engine and another on the Digital Elevation Model (DEM) dataset.

A secondary contribution covers the implementation of a pathfinding engine that utilizes open and crowd-sourced data from the OpenStreetMap project as well as SRTM elevation data. The overview demonstrates the ability to conduct additional research with data that is free, accurate and constantly improving.

## 2.0 LITERATURE REVIEW

The literature review is broken up into two sections. Section 2.1 provides a cursory overview of research into DEFs that this research builds upon. Section 2.2 provides a review of research into fuel-efficient routing, fuel consumption simulation, and the application of DEMs.

### 2.1 Circuitry factors and distance estimating functions

DEFs are commonly used in logistics applications such as facility location, fleet sizing and network design. Ballou et al. (Ballou, et al., 2002) provided a summary of research and calculated circuitry factors for long-distance paths in various countries and provided good results with the following equation

$$D_{ab} \approx b_0 \cos^{-1}(\sin(y_a) \sin(y_b) + \cos(y_a) \cos(y_b) \cos(|y_b - y_a|)) \quad (1)$$

where  $x$  and  $y$  are respectively the longitude and latitude coordinates of the origin and destination. Equation (1) represents the great-circle distance, which is often needed for long-distance calculations where Euclidean distance calculation will produce significant errors.

Goncalvesa et al. (Gonçalvesa, et al., 2014) evaluated the estimation of circuitry factors in Brazilian soy bean supply chains. The authors tested a non-zero intercept regression formula using the Euclidean distance:

$$D_{ab} \approx b_0 + b_1 \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \quad (2)$$

The authors also discussed limitations for shorter distance estimation and when the constant is less than zero.

Shihad et al. (Shahid, et al., 2009) applied a DEF to estimating distances of patients to a hospital in the context of the classic facility location problem. The authors tested regression functions that included a Manhattan distance metric as well as the more generalized Minkowski formula

$$D_{ab} \approx [(x_b - x_a)^p + (y_b - y_a)^p]^{\frac{1}{p}}, p > 0 \in \mathbb{N} \quad (3)$$

where the variable  $p$  is estimated in the interval  $1 \leq p \leq 2$ ;  $p = 1$  represents a pure Manhattan distance and  $p = 2$  the Euclidean distance. The authors found a value of  $p = 1.52$  to provide the best fit on their dataset for travel distance (roughly equal contribution of Manhattan and Euclidean distance) but found a lower value (1.23) best fits travel time calculations.

To the best of this author's knowledge, no generalized approach to fuel-consumption estimating functions has yet been proposed in the existing literature.

## **2.2 Fuel-efficient routing and modeling**

Fuel-efficient pathfinding has been implemented in several commercial software applications. Schaper and Bruns (Schaper & Bruns, 2015) presented an implementation of a route calculation engine (RCE) for fuel-efficient routing of commercial vehicles with detailed discussion of data processing and algorithmic design.

Simulation of vehicle fuel-consumption has also garnered much research. An important factor for simulations is the accuracy of the DEM. Wood et al. (Wood, et al., 2014) developed methods for processing the USGS 1/3 arc-second resolution elevation data and applied it to a road network. Their methods focused on filling gaps and correcting inaccuracies caused by bridges and tree canopies. The authors applied their dataset to a fuel simulation model (Wood, et al., 2014) and deduced that grade alone is responsible for about 1-3% of commercial vehicle fuel consumption for long-distance routes.

Much of the existing literature on the estimation of vehicle fuel consumption has focused on measuring the relationship between vehicle-specific variables and traffic congestion. Barth et al. (Barth, et al., 2005) provided a comprehensive review of research into fuel consumption and emissions modeling for heavy duty diesel (HDD) vehicles. The authors also validated a commonly used empirical formula for fuel consumption based on travel speed that is used in this paper.

Cappiello et al. (Cappiello, et al., 2002) developed statistical models to predict fuel-consumption and emissions based on vehicle characteristics and load. More recently Wyatt et al. (Wyatt, et al., 2014) developed statistical models of fuel consumption and CO<sub>2</sub> emissions as a function of grade. Lopp et al. (Lopp, et al., 2015) conducted similar research for commercial vehicles and develop percentage increase factors for fuel consumption.

To this author's knowledge, no research has yet been done that generalizes calculations based on topographical properties or travel time characteristics of an urban area.

### 3.0 METHODOLOGY

This section covers the data sources utilized, as well as an overview of the classic shortest path problem. Section 3.1 provides details of the OSM and SRTM data sources and their usage. Section 3.2 covers the formulation of the shortest path problem with a brief discussion of optimization techniques. Section 3.3 concludes with an overview of the data processor and pathfinding engine.

#### 3.1 Data Sources

##### 3.1.1 *OpenStreetMap*

OSM provides raw data in three different formats: XML (OpenStreetMap Wiki contributors, 2017); Protocol buffer Binary Format (PBF), a compressed format developed in part by Google (OpenStreetMap Wiki contributors, 2018); and OSM (OSM XML 5 times smaller) (OpenStreetMap Wiki contributors, 2018), which has a similar hierarchical structure to the XML format but with compressed data types. The XML format is chosen for simplicity of data manipulation and for

An OSM dataset consists of the fundamental elements nodes, ways and relations. Each element is assigned a unique 64-bit signed integer identifier and contains a set of one or more key-value pairs of data describing various features. Subcategories are denoted with colons to denote more specific information (e.g. Relation: Restriction for turn restrictions). Figure 1, Figure 2, and Figure 3 provide excerpts of the XML raw data for ways, nodes, and relations, respectively.

```
<way id="5374116" user="balrog-kun" uid="20587" visible="true" version="2" changeset="4191243" timestamp="2010-03-21T17:00:41Z">
  <nd ref="38474796"/>
  <nd ref="38532325"/>
  <nd ref="38512569"/>
  <nd ref="38532328"/>
  <tag k="highway" v="residential"/>
  <tag k="name" v="Southwest 18th Street"/>
  <tag k="tiger:cfcc" v="A41"/>
  <tag k="tiger:county" v="Deschutes, OR"/>
  <tag k="tiger:name_base" v="18th"/>
  <tag k="tiger:name_direction_prefix" v="SW"/>
  <tag k="tiger:name_type" v="St"/>
  <tag k="tiger:reviewed" v="no"/>
  <tag k="tiger:separated" v="no"/>
  <tag k="tiger:source" v="tiger_import_dch_v0.6 20070808"/>
  <tag k="tiger:tlid" v="152547309:152547311:152562230"/>
  <tag k="tiger:zip_left" v="97756"/>
  <tag k="tiger:zip_right" v="97756"/>
</way>
```

Figure 1: XML for OpenStreetMap way



```

<node id="1804147365" lat="44.2704118" lon="-121.2113090" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147367" lat="44.2704199" lon="-121.2115181" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147369" lat="44.2704221" lon="-121.2111109" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147371" lat="44.2704268" lon="-121.2100871" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147373" lat="44.2705371" lon="-121.2098846" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147374" lat="44.2705371" lon="-121.2100871" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147375" lat="44.2705371" lon="-121.2102945" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147376" lat="44.2705371" lon="-121.2105019" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147377" lat="44.2705371" lon="-121.2107019" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147378" lat="44.2705371" lon="-121.2109044" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147379" lat="44.2705371" lon="-121.2111217" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147380" lat="44.2705371" lon="-121.2113143" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:20Z"/>
<node id="1804147381" lat="44.2705371" lon="-121.2115242" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<node id="1804147382" lat="44.2709457" lon="-121.2104352" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<node id="1804147383" lat="44.2709507" lon="-121.2100860" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<node id="1804147384" lat="44.2710319" lon="-121.2100842" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<node id="1804147396" lat="44.2719312" lon="-121.2115878" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<tag k="power" v="tower"/>
</node>
<node id="1804147400" lat="44.2720108" lon="-121.2111112" user="zephyr" uid="121241" visible="true" version="1" changeset="12038888" timestamp="2012-06-27T17:17:21Z"/>
<tag k="power" v="tower"/>
</node>

```

Figure 2: XML for a block of OpenStreetMap nodes

```

<relation id="1946180" user="Paul Johnson" uid="92286" visible="true" version="1" changeset="10258474" timestamp="2011-12-31T23:00:29Z">
  <member type="way" ref="143657625" role="from"/>
  <member type="way" ref="143657578" role="to"/>
  <member type="node" ref="1572175029" role="via"/>
  <tag k="restriction" v="only_right_turn"/>
  <tag k="type" v="restriction"/>
</relation>

```

Figure 3: XML for OpenStreetMap relation (type restriction)

Within the *processed* dataset elements are referred to as edges, vertices and shape points (Figure 4). Vertices represent decision points, edges represent connections between vertices (and their associated costs), and shape points provide rendering and elevation data for each edge.

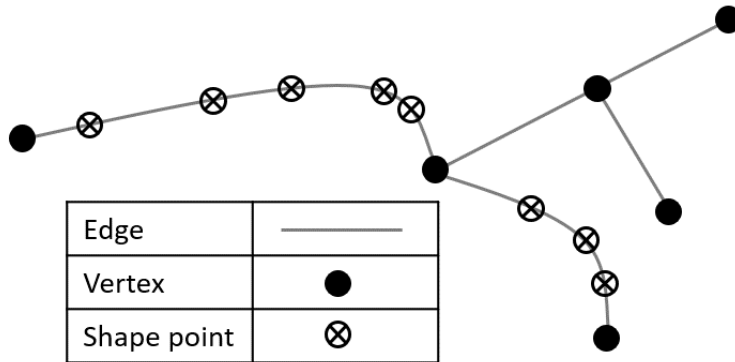





Figure 4: Key graph elements

The OSM nodes and ways are processed to identify intersection points and generate the edges, vertices and initial set of shape points. Relations of type restriction are used to define allowed travel between adjacent edges in the graph. The data provided in the restriction type of a relation define either prohibited or allowed movements, and optionally by vehicle type. A summary of data usage is provided in Table 1.

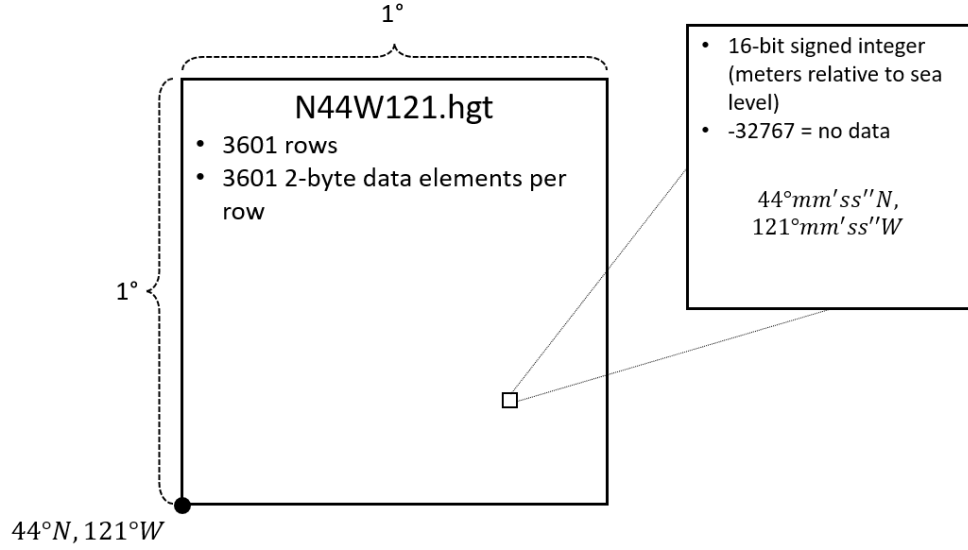
**Table 1: OSM data elements and usage.**

Data Element	Attributes	Usage
Node 	<ul style="list-style-type: none"> <li>Latitude/longitude coordinates</li> <li>POI features</li> <li>Barriers/obstructions</li> </ul>	<ul style="list-style-type: none"> <li>Graph vertices</li> <li>Shape points</li> <li>Address information and geocoding</li> <li>Calculation of turn movements (natural language directions and turn costs)</li> </ul>
Way 	<ul style="list-style-type: none"> <li>Road category</li> <li>Permitted/prohibited vehicle types</li> <li>Road geometry</li> </ul>	<ul style="list-style-type: none"> <li>Graph edges</li> <li>Vehicle access restrictions</li> <li>Hierarchical routing (speed-up techniques)</li> <li>Route rendering</li> </ul>
Relation 	<ul style="list-style-type: none"> <li>Allowed/Prohibited movements between ways</li> <li>Allow/prohibited vehicle types</li> </ul>	<ul style="list-style-type: none"> <li>Turn restrictions</li> </ul>

### 3.1.2 Shuttle Radar Topography Mission

The elevation data come from NASA's SRTM dataset (USGS, 2018). The original world-wide dataset had a resolution of 3 arc seconds; however, in 2014 NASA released a new 1-arc second dataset with worldwide coverage.

SRTM data files are organized into 1x1 degree grid sections with the file name referencing the lower left corner of the grid. The data are arranged in 3,601 rows of 3,601 samples represented as signed 16-bit integers indicating the elevation relative to sea level in meters. Values of -32768 indicate no data for the coordinate location. Figure 5 provides a visual representation of a single SRTM .hgt file. The elevation coordinates are applied to the OSM dataset using a simple linear interpolation method to assign values to existing shape points and create additional ones.



**Figure 5: SRTM file format.**

### 3.2 Shortest Path Problem Statement

The shortest path problem is formulated on a directed graph  $G(V, E)$  with a set of vertices  $V$  connected by a set of edges  $E$ . For every edge  $e_{ij} \in E$  a cost  $c$  exists such that

$$c(e_{ij}) = c(v_i, v_j) = \begin{cases} \mathbb{R}_0^+ & \text{Cost must be } > 0 \\ \infty & \text{Vertices } v_i \text{ and } v_j \text{ are not connected} \end{cases} \quad (4)$$

Although  $c(e_{ij}) = c(e_{ji})$  for shortest paths, this is not strictly the case for paths that minimize travel time, and generally not the case for fuel-efficient routes.

A path  $P_{st}$  from an origin  $v_s$  to a destination  $v_t$  is a sequence of vertices  $v_s = v_1, v_2, v_3 \dots, v_p = v_t$ . The total cost of a path from  $v_s$  to  $v_t$  can be expressed as

$$c(P_{st}) = \sum_{i=1}^{p-1} c(v_i, v_{i+1}). \quad (5)$$

The objective is then to minimize the path cost such that

$$\delta(v_s, v_t) = \begin{cases} \min(c(R_{s,t})) & \text{if a path from } s \text{ to } t \text{ exists} \\ \infty & \text{No path exists} \end{cases} \quad (6)$$

### 3.2.1 Dijkstra's Algorithm

Although it is possible to formulate the shortest path problem as a linear or dynamic programming problem, it is much more efficient to solve using an iterative approach known as Dijkstra's algorithm. The algorithm works as follows:

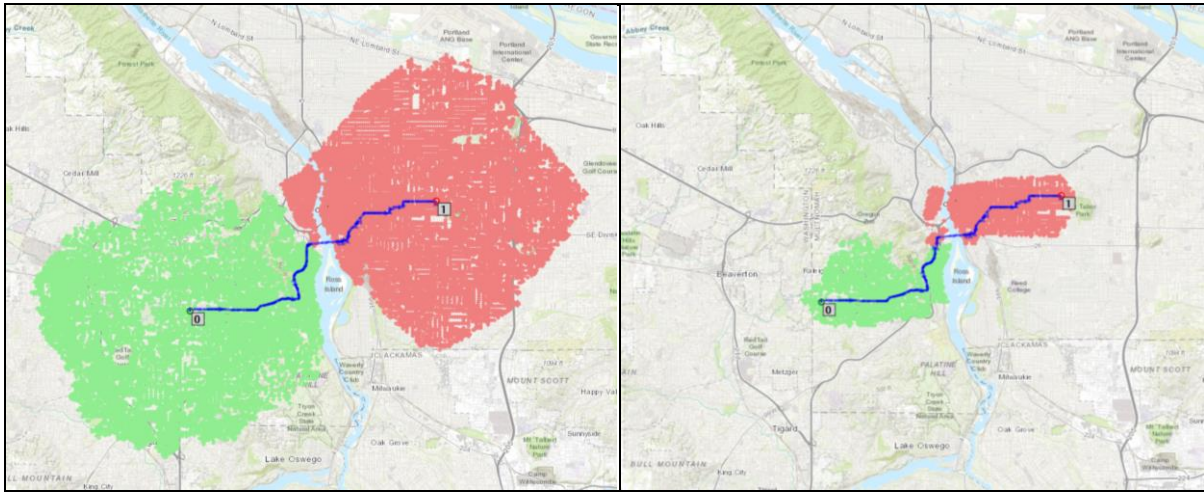
1. Initialize the source vertex  $v_s$  with cost 0 and all other vertices to  $\infty$
2. While there are unvisited vertices, select the path (denoted  $P_i$ ) with minimum cumulative cost.
3. Mark the vertex (denoted  $v_i$ ) as visited.
4. Calculate the cost for all unvisited neighbor vertices  $v_j$  reachable from  $v_i$  by appending their cost  $c(v_i, v_j)$  to the total cost of  $P_i$  and store each path sorted by cumulative cost.
5. Stop when the end vertex of the path selected in step 2 is  $v_t$  or no more paths can be extracted (no solution).

For a given input graph, the performance of Dijkstra's algorithm is highly dependent on the data structures used to store visited nodes and the partial path trees sorted by their cumulative costs. Visited vertices are often stored in a *closed set* implemented as a hash table with amortized  $O(1)$  insert and search complexity. The partial paths are stored in a heap or priority queue with  $O(\log n)$  insert and removal where  $n$  is the number of partial paths.

### 3.2.2 A\* Algorithm

The A\* algorithm is an extension of Dijkstra's algorithm and adds a heuristic that estimates the cost to  $v_t$  which is stored with each partial path. The result is a more directed search that selects the lowest cost partial path at each iteration based on an estimated total cost to  $v_t$ . The estimating

heuristic is considered *admissible* if it never overestimates the cost to reach  $v_t$ , which guarantees the algorithm will find the optimal path. For shortest path routing this is often calculated as the Euclidean distance from the end of each candidate path to  $v_t$  since no path can ever be shorter than the Euclidean distance from  $v_s$  to  $v_t$ . In general, the closer the heuristic is to estimating the actual remaining cost to reach  $v_t$  without overestimating, the more efficient the algorithm will run with fewer iterations before reaching  $v_t$ . Figure 6 shows a shortest path calculation using Dijkstra's algorithm and A\* (implemented as bi-directional). Knowing only the cost from the source vertex, Dijkstra's algorithm explores in a roughly circular pattern in all directions until the destination is reached; in comparison the A\* search explores nodes in a more focused beam shape towards the destination.



**Figure 6: Shortest path calculations with bi-directional Dijkstra's algorithm (left) and A\* (right) with a Euclidean distance heuristic. Green dots indicate vertices explored by the forward search; red dots indicate those explored by the backwards search.**

### 3.2.3 Bi-directional algorithms

Both Dijkstra's algorithm and A\* can be formulated as bi-directional algorithms, with alternating searches from  $v_s$  to  $v_t$  and vice versa. The search terminates when the minimum-cost partial path extracted by one search terminates at a vertex visited by the other search. The optimal path is then constructed from the minimum-cost partial paths generating by each search. Optimality is guaranteed if the algorithm alternates between the forward and reverse searches in a serial manner. In all cases performance is enhanced at the cost of slightly larger overhead to maintain two sets of open and closed lists.

### 3.2.4 *A\* performance enhancements*

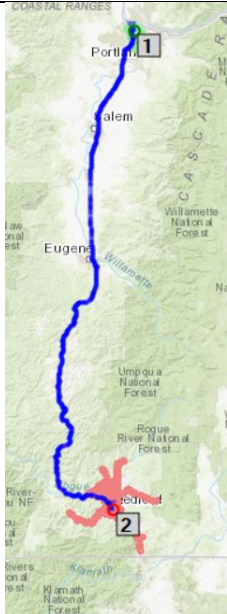

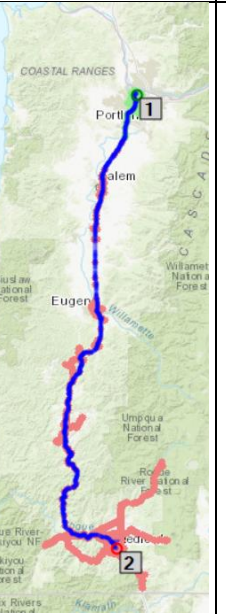

A great deal of research has been dedicated to improving the performance of A\* in road networks. Although a simple Euclidean distance heuristic adds some performance improvement over Dijkstra’s algorithm, it is often insignificant in large continental-sized road networks. The lower bound established by the heuristic must often be scaled up substantially to obtain acceptable improvements in runtime (Goldberg & Harrelson, 2005).

Most speed-up techniques involve some preprocessing of the graph inclusive of pre-determined edge costs. The hierarchical nature of road networks allows for several approaches for compressing or simplifying the graph. Long distance paths usually involve a small amount of navigation on local streets, then exclusively use higher-speed freeways and major highways. An approach proposed by Pfoser et al. (Pfoser, et al., 2009) uses the natural road hierarchy embedded in the map data to prune the search space without preprocessing. The algorithm does not guarantee optimality but is useful for illustrating the efficiency improvements of hierarchical techniques. At each iteration of the bi-directional search the allowed vertices from a candidate path must be accessible from edges that are of an equal or more important road category. The best category is stored with the path, so each successive candidate path progressively moves to a higher category and “sees” fewer available nodes to expand. The algorithm alternates between the forward and reverse searches such that both are maintained on the same road category until they meet.

Table 2 shows results for a path run with 3 different levels of hierarchical pruning. A common benchmark that is hardware and software-independent is the ratio of vertices on the resultant shortest path divided by the total number explored (contained in the closed set).

More recent techniques require preprocessing but guarantee optimality and reduce the storage space of the graph via recursive compression and the creation of shortcuts (Geisberger, et al., 2008), a method known as *contraction hierarchies*. This approach requires techniques for efficiently updating the graph when edge costs change (i.e. dynamic travel times, road closures, etc.), which has also been an active area of research (Abraham, et al., 2010). Together these techniques are the predominant methods of calculating exact shortest paths with the least computation time.

**Table 2: Comparison of hierarchical optimizations (HBA\*) vs. regular bi-directional A\***

	Hierarchical Bi-directional A* (HBA*) algorithm			A* Euclidean
Explored vertices	10,034	11,996	14,892	117,611
				

### 3.3 Pathfinding Engine Overview

The pathfinding engine is comprised of a data processor that imports the OSM XML files and SRTM binary files. The OSM file is processed first to create the graph data structure and process the road attributes as covered in section 3.1. Figure 7 provides a summary of the data processing steps.

After processing the OSM file, the SRTM data files are applied to produce a digital elevation model (DEM). Coordinates in the SRTM files with corresponding elevation values are linearly interpolated onto the graph edges.

The pathfinding engine supports batched many-to-many routing calculations and stores the results in the SQL database for analysis. The raw SRTM data are also stored separate from the shape point data in a spatially indexed table for analysis covered in section 4.2.3.

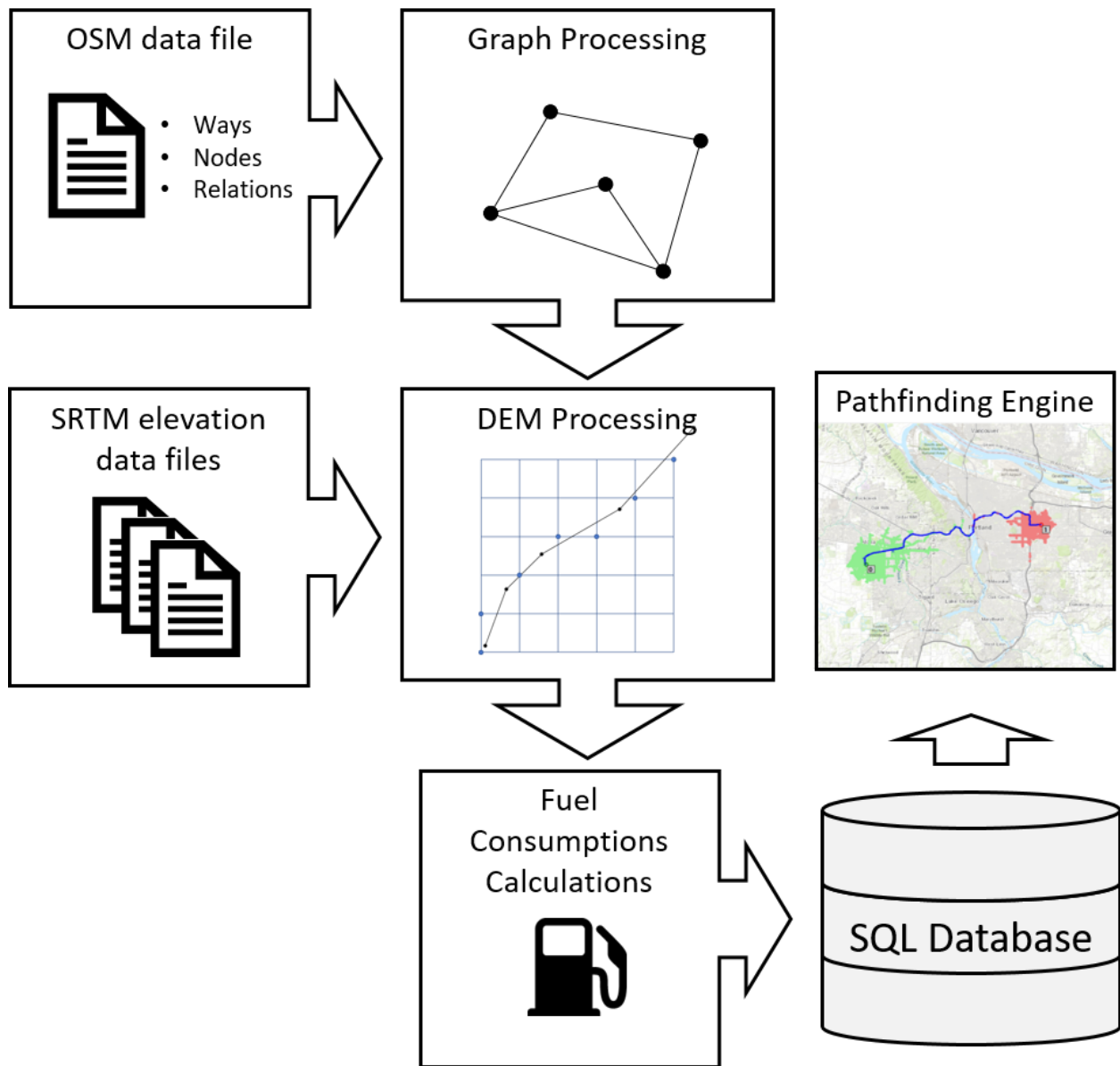


Figure 7: Pathfinding engine data processing



## 4.0 ANALYSIS

### 4.1 Experimental Setting

#### 4.1.1 Fuel consumption calculation

The fuel consumption cost  $f_{ij}$  of traversing each edge  $e_{ij}$  in the graph is adapted from Franceschetti et al (Franceschetti, et al., 2013) and given by

$$f_{ij} = \sum_{k=1}^{n_{ij}-1} \frac{d(p_{ij,k}, p_{ij,k+1})}{\kappa\psi} \max \left[ \frac{kN_e V}{v_{ij}} + \frac{\frac{1}{2} C_d \rho A v_{ij}^2 + \mu g (\sin(\tan^{-1} \sigma_{ij,k,k+1}) + C_r \cos(\tan^{-1} \sigma_{ij,k,k+1}))}{1000 \varepsilon \omega}, 0.001 \right], \quad (7)$$

$$\sigma_{ij,k,k+1} \equiv \frac{z_{ij,k+1} - z_{ij,k}}{d(p_{ij,k}, p_{ij,k+1})}$$

where  $p_{ij,k}$  is the  $k$ th shape point on edge  $e_{ij}$ ;  $z_{ij,k}$  is the elevation in meters of shape point  $p_{ij,k}$ ;  $n_{ij}$  is the total number of shape points on edge  $e_{ij}$ ;  $d(\cdot, \cdot)$  is the distance between adjacent shape points. The remaining constants in equation (7) are defined in Table 3. The original equation proposed by Franceschetti et al (2013) had to be adjusted to prevent negative fuel cost values when traveling on a sufficiently steep downgrade. To this author's knowledge, there are no general guidelines for handling this situation and the amount of fuel consumed when the vehicle is essentially idling varies widely based on engine characteristics. Figure 8 shows fuel consumption as a function of travel speed for different road grades utilizing the parameters contained in Table 3.

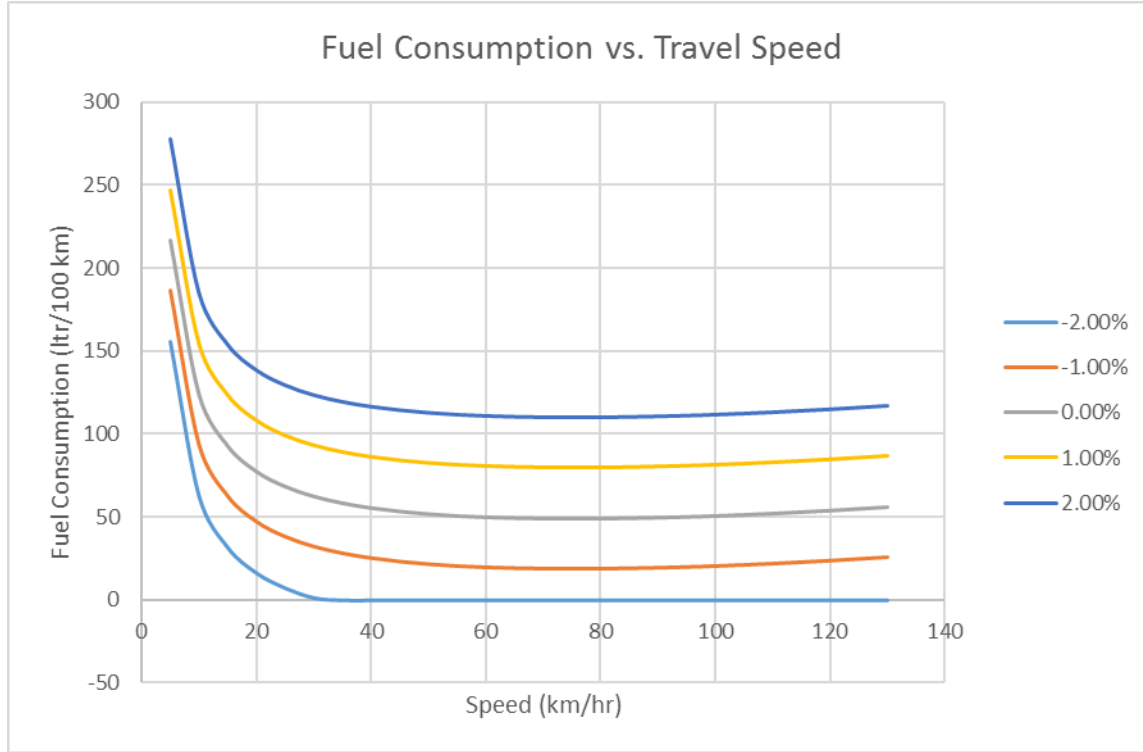


Figure 8: Fuel consumed as a function of travel speed at different road grades

Table 3: Vehicle fuel consumption parameters and values

Variable	Definition	Value
$\kappa$	Heating value of typical diesel fuel (kJ/g)	44
$\psi$	Conversion factor grams to liters	737
$k$	Engine friction factor (kJ/rev/l)	0.2
$N_e$	Engine speed (rev/s)	33
$V$	Engine displacement (l)	12.9
$\rho$	Air density (kg/m <sup>3</sup> )	1.2041
$A$	Frontal surface area (m <sup>2</sup> )	3.912
$\mu$	Vehicle weight (kg)	Variable (6350-36400)
$g$	Gravitational constant (m/s <sup>2</sup> )	9.81
$C_d$	Coefficient of aerodynamic drag	0.7
$C_r$	Coefficient of rolling resistance	0.01
$\epsilon$	Vehicle drive train efficiency	0.4
$\varpi$	Efficiency parameter of diesel engines	0.9

Equation (7) is calculated with constant travel speed  $v_{ij}$  over each edge  $e_{ij}$ . Using the OSM highway flag, speeds are calculated according to Table 4. While actual travel time data is much more desirable compared to statically calculated values, an enormous amount of data would have to be obtained for all edges in the graph to obtain consistent results from the pathfinding engine. Such an endeavor is outside the scope of this research, but certainly worthy of future investigation.

**Table 4: Travel speeds calculated by OpenStreetMap category (highway tag)**

Category	Description	Travel Speed (km/hr.)
Motorway	Restricted access, high-speed	120
Trunk	Highest-level road class without restricted access; typically US routes in the United States	95
Primary	Major highway linking large towns	75
Secondary	Major urban road and arterials	55
Tertiary	Arterial connectors and collectors	45
Residential	Neighborhood and local-access only	35

#### 4.1.2 Characteristics of Fuel-efficient Paths

It is worthwhile to assess the characteristics of shortest, fastest and fuel-efficient paths to determine tradeoffs between the different objectives. Table 5 compares the key characteristics of the three optimization strategies. It is notable that the fuel-efficient path only sacrifices a modest increase in distance and travel time compared to the potential savings in fuel over a shortest path. This is an expected result as both distance and travel speed impact fuel, with travel speed having a large impact on fuel-consumption as it approaches traffic congestion speeds. Therefore, it is intuitive that fuel-efficient paths strike a balance between shortest and fastest.

It is evident that fuel-efficient paths compete most directly with fastest paths as both sacrifice about equal amounts of increased travel distance over a pure shortest path. Fuel-efficient paths give up slightly less travel time compared to the extra fuel of fastest paths. Thus, in cases where marginal costs of travel time and fuel are roughly equal, a fuel-efficient strategy would be the preferred choice for a commercial operator.

**Table 5: Comparison of key metrics for the three path optimization strategies**

Path Type	Avg. Fuel Consumption (ltr.)	Increase	Path Type	Avg. Distance (km)	Increase	Path Type	Avg. Travel Time (min.)	Increase
Fuel-efficient	17.22	-	Shortest	18.41	-	Fastest	15.10	-
Fastest	18.59	7.38%	Fuel-efficient	20.20	8.88%	Fuel-efficient	15.99	5.61%
Shortest	22.69	24.11%	Fastest	20.33	9.47%	Shortest	19.16	21.23%

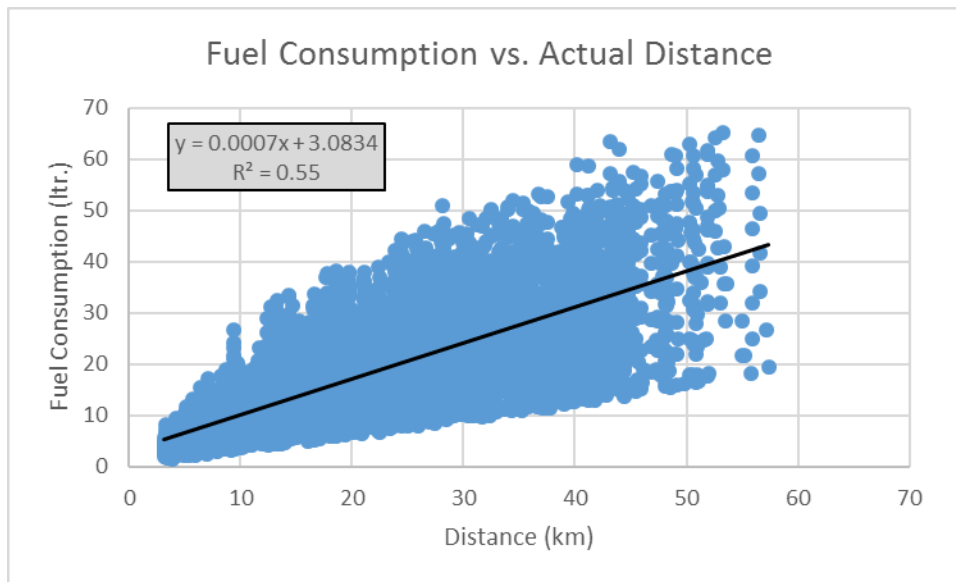
## 4.2 Fuel Consumption Estimation Models

Regression models are formulated directly from equation (7) with the independent variables derived as described in the following sections. Section 4.2.1 provides an overview of the statistical

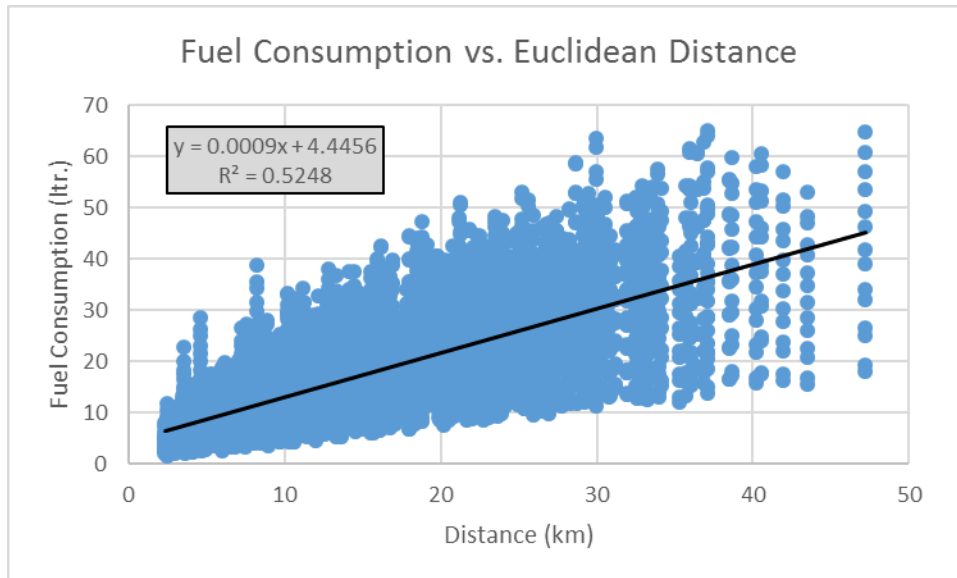
techniques utilized for the model development and well as justification for the FCEF approach compared to a simple DEF. Section 4.2.2 covers variables derived from the graph dataset used for routing. Section 4.2.3 concludes with models utilizing only the DEM dataset and GIS query techniques available in Microsoft SQL Server.

#### 4.2.1 Initial Development

Before developing the statistical models, the relationship between travel distance and fuel consumption is examined to determine the overall suitability of a DEF for predicting fuel consumption. Figure 9 shows the relationship between fuel consumption and actual route distance. The modest R-squared value and even distribution about the regression line suggest that additional independent variables need to be included to accurately predict fuel consumption. This is confirmed in Figure 10 with fuel consumption vs. Euclidean distance with a similar R-squared. These results suggest traditional DEFs are inadequate for estimating fuel consumption.

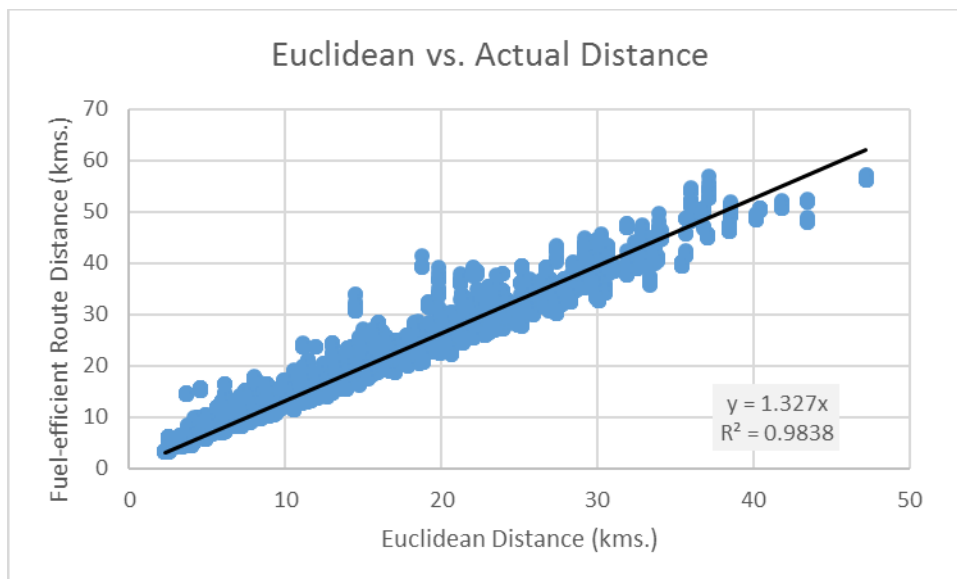


**Figure 9: Fuel consumption vs. actual travel distance**



**Figure 10: Fuel consumption vs. Euclidean distance**

Despite the relatively weak relationship between distance and fuel consumption, the travel distance of fuel-efficient routes still tracks closely with the Euclidean distance as shown in Figure 11. This suggests the Euclidean distance is still an important basis upon which to build the FCEF models upon. This also suggests that fuel-efficient paths follow similar trends as shortest distance and travel time from existing research with real-world applications (Figliozzi, 2008).



**Figure 11: Euclidean vs. Actual Travel Distance for fuel-efficient paths**

The statistical models developed in sections 4.2.2 and 4.2.3 are derived directly from equation (7). It can first be observed that equation (7) consists of a speed term and an elevation term. The goal is to build up the models incrementally from a simple DEF to determine how effective the statistical predictors alone can account for parameters that are more difficult to measure, namely expected travel speed and elevation profile of a path for a given origin-destination pair.

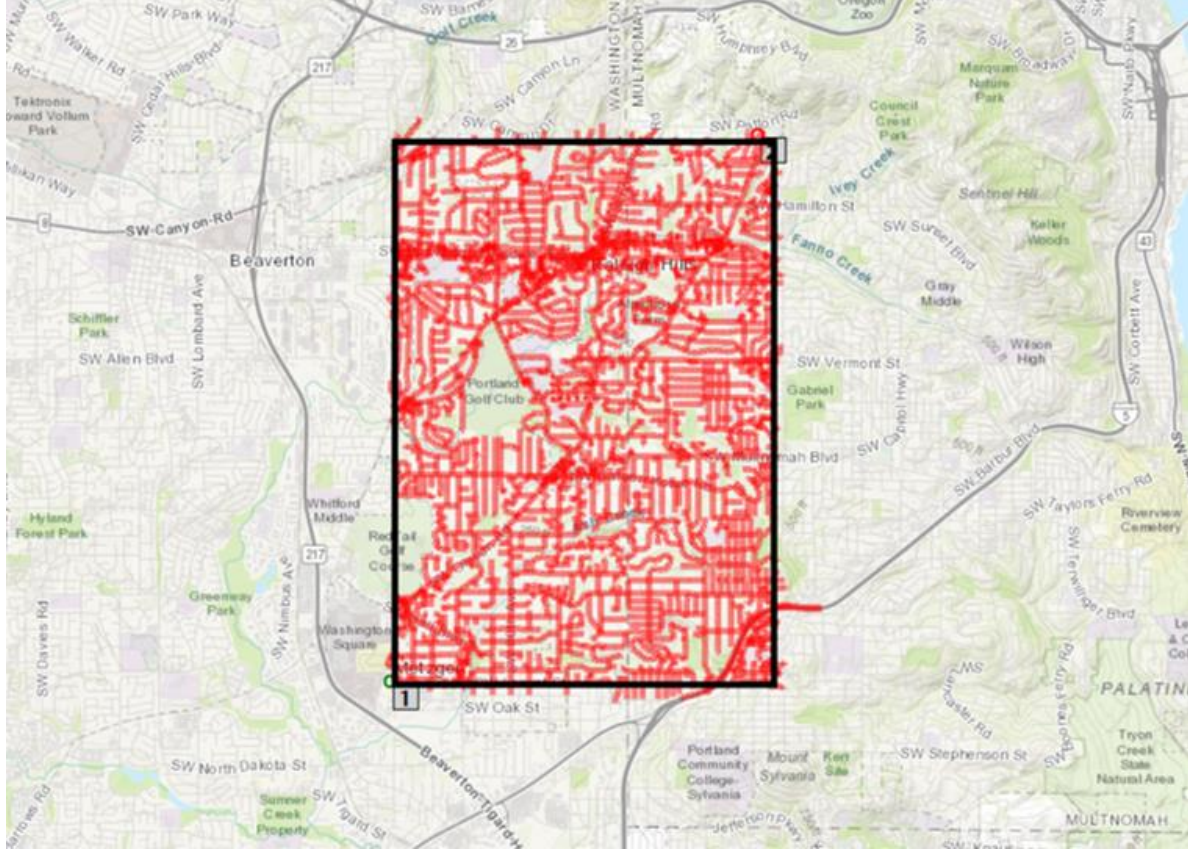
The estimation of travel speed is discussed first. Using speed values from Table 4, the average travel speeds are calculated for the aggregated data by urban test area and by path. First let  $G(V, E)$  represent the graph for an urban area with subgraphs  $G_{st}(V, E) \subset G(V, E)$  for each origin-destination pair  $v_s$  and  $v_t$ .  $G_{st}(V, E)$  is defined as the set of edges contained in a bounding box enclosing  $v_s$  and  $v_t$  (Figure 12). The practical implementation requires spatially indexing the edges to guarantee a fast calculation, a trivial task in most relational database systems.

The average travel speed in these defined networks is calculated using the travel speed along all edges in  $G(V, E)$  and  $G_{st}(V, E)$  weighted by the edge distance such that

$$\bar{v} = \frac{\sum_{\forall e_{ij} \in G(V, E)} d_{ij} v_{ij}}{\sum_{\forall e_{ij} \in G(V, E)} d_{ij}}, \quad (8)$$

$$\bar{v}_{st} = \frac{\sum_{\forall e_{ij} \in G_{st}(V, E)} d_{ij} v_{ij}}{\sum_{\forall e_{ij} \in G_{st}(V, E)} d_{ij}} \quad (9)$$

Where  $d_{ij}$  and  $v_{ij}$  are the distance and travel time along an edge  $e_{ij}$ , respectively.



**Figure 12: Bounding box for a route with queried edges used to calculate  $v_{st}$  and  $\alpha_{st}$ .**

Similar methodology is employed to capture the impact of elevation by introducing the concept of an average absolute elevation change. From the terminology provided in Figure 4, a given ordered shape point on an edge  $e_{ij}$  is denoted  $p_{ij,k}$   $k \in \mathbb{N}_1[1 \dots n_{ij}]$  with corresponding elevation  $z_{ij,k}$  and  $n_{ij}$  denoting the number of shape points on  $e_{ij}$ . The average change in elevation is calculated as

$$\bar{\alpha} = \frac{\sum_{e_{ij} \in G(V,E)} \left[ \sum_k^{n_{ij}-1} \frac{|z_{ij,k+1} - z_{ij,k}|}{d(p_{ij,k}, p_{ij,k+1})} \right]}{\sum_{e_{ij} \in G(V,E)} (n_{ij} - 1)}, \quad (10)$$

$$\bar{\alpha}_{st} = \frac{\sum_{e_{ij} \in G_{st}(V,E)} \left[ \sum_k^{n_{ij}-1} \frac{|z_{ij,k+1} - z_{ij,k}|}{d(p_{ij,k}, p_{ij,k+1})} \right]}{\sum_{e_{ij} \in G_{st}(V,E)} (n_{ij} - 1)} \quad (11)$$

where  $d(\cdot, \cdot)$  is the distance between adjacent shape points. As with equations (8) and (9), equations (10) and (11) calculate these measures for the urban test area and at the path level, respectively.

The absolute value of the elevation difference is utilized based on the conjecture that increasing changes in elevation will increase fuel consumption, and a net negative change in elevation will not produce any appreciable reduction in fuel consumption. This has the desirable property for paths on flat terrain in that the contribution to fuel consumption is zero. Table 6 provides a summary of the regression parameters and units.

**Table 6: Regression estimators and units**

Estimator	Interpretation	Units
$b_v$	Travel speed estimator	Dimensionless
$b_u$	Vehicle mass estimator	Dimensionless
$b_\alpha, b_\lambda$	Increase per unit elevation change	Meters

#### 4.2.2 Graph dataset models

Using equations (8), (9), (10), and (11), regression models are derived and given in Table 7 with  $r_{st}$  the Euclidean distance between  $v_s$  and  $v_t$ ,  $\mu$  the mass of the vehicle, and  $(b_v, b_\mu, b_\alpha)$  the set of parameters to estimate via linear regression. For brevity, values that are constant in the regressions are defined as follows:

$$\gamma \equiv \frac{kN_e V}{\kappa\psi},$$

$$\xi \equiv \frac{C_d \rho A}{1000 \kappa \psi \varepsilon \omega},$$

$$\eta \equiv \frac{g}{1000 \kappa \psi \varepsilon \omega}.$$

**Table 7: Graph-based regression models**

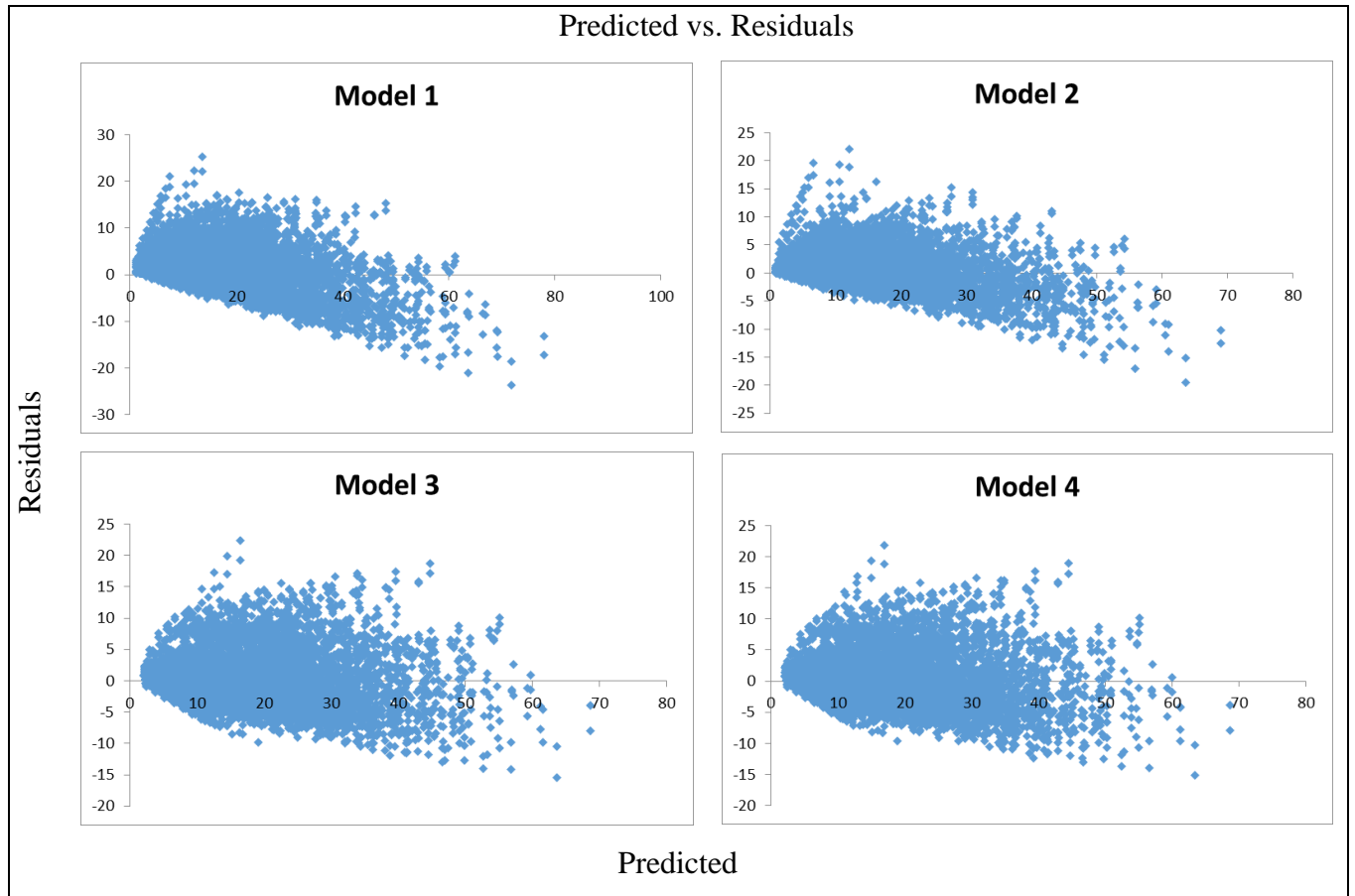
Model 1	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}} + \xi \bar{v}^2 \right) + b_\mu r_{st} \mu \eta$
Model 2	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}_{st}} + \xi \bar{v}_{st}^2 \right) + b_\mu r_{st} \mu \eta$
Model 3	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}} + \xi \bar{v}^2 \right) + b_\mu r_{st} \mu \eta + b_\alpha \bar{\alpha} \mu \eta$
Model 4	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}_{st}} + \xi \bar{v}_{st}^2 \right) + b_\mu r_{st} \mu \eta + b_\alpha \bar{\alpha} \mu \eta$



Model 1 and Model 2 are based only on attributes of the vehicle and the estimated average travel speed of the road network. The parameters  $b_v$  and  $b_\mu$  are both dimensionless, with the latter estimating the contribution of vehicle mass changes to the total fuel consumption.

Model 3 and Model 4 introduce the average absolute grade dimensionless independent variables  $\bar{\alpha}$  and  $\bar{\alpha}_{st}$  from equations (10) and (11) that attempt to capture the impact of elevation change by test area and path, respectively. The estimator  $b_\mu$  remains dimensionless and independently estimates the effect of vehicle mass on the result. The added estimator  $b_\alpha$  has units of distance (meters) and can be interpreted as the increase in fuel consumption per unit change in elevation. The introduction of the straight-line (Euclidian) distance  $r_{st}$  into the third term of these models produced poor statistical significance and suggested a collinearity between the second and third terms. The best results were obtained by separately accounting for the combinations of distance/vehicle mass and elevation change/vehicle mass.

Figure 13 provides plots of predicted vs. residuals for the four models. The plots for Model 3 and Model 4 suggest that including the independent variables capturing elevation change may improve the quality of the regression results as residuals are more evenly distributed about the x-axis.



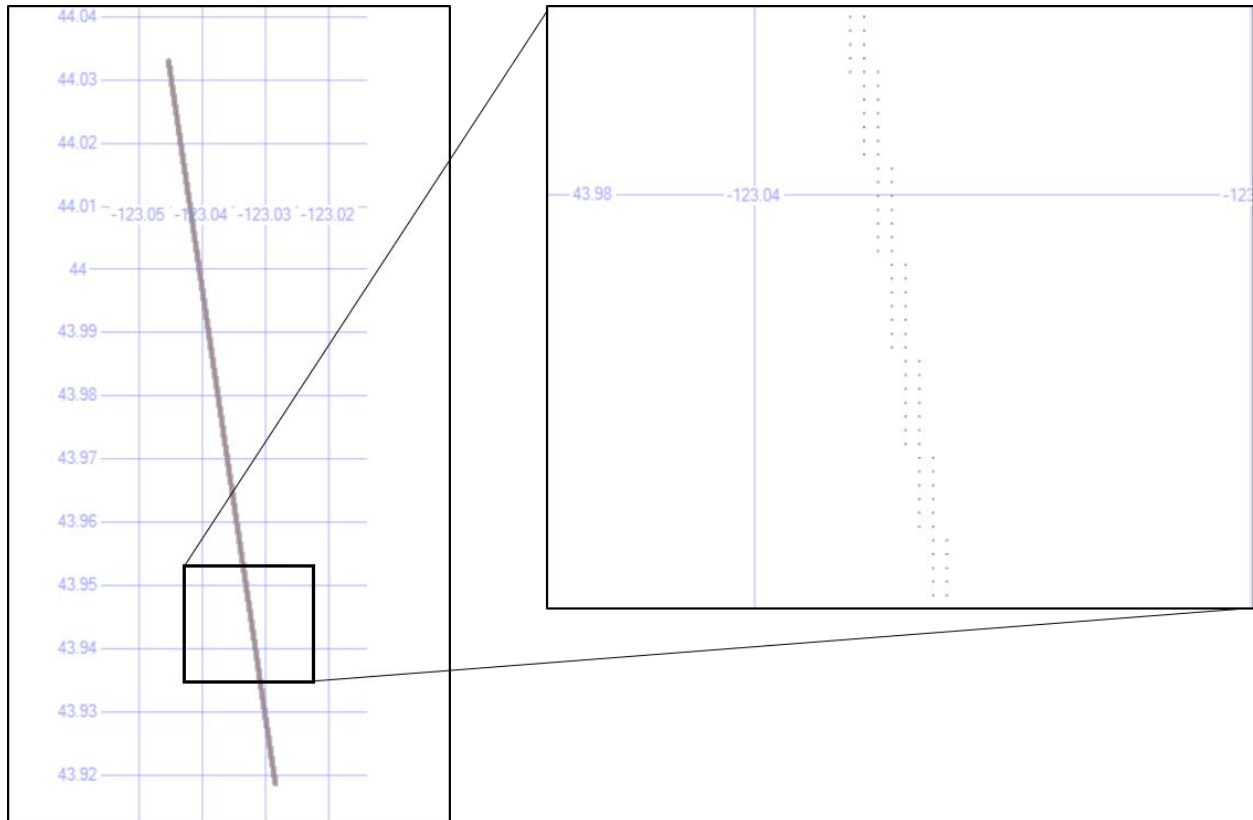
**Figure 13: Predicted vs. residual plots for Models 1-4**

#### 4.2.3 SRTM dataset models

Two additional models are tested on the SRTM DEM dataset as alternatives to models 3 and 4. The assumption in a theoretical planning context is that a complete elevation dataset based on the pathfinding engine's graph is unavailable. The goal is to substitute measures of elevation variance derived from the SRTM data as a proxy for the elevation profile along the expected path.

The SRTM data are loaded into SQL Server for each of the urban areas tested and spatially indexed by coordinate. For each route tested a line string geography object is constructed between the origin and destination and saved in the database. The line string is then inflated by a factor equal

to the resolution of the SRTM data points (1/3600 degree) using STBuffer<sup>1</sup> to create a polygonal geometry. Figure 14 shows images from SQL Server Management Studio's spatial viewer, with the constructed line string on the left and intersecting SRTM coordinates on the right.

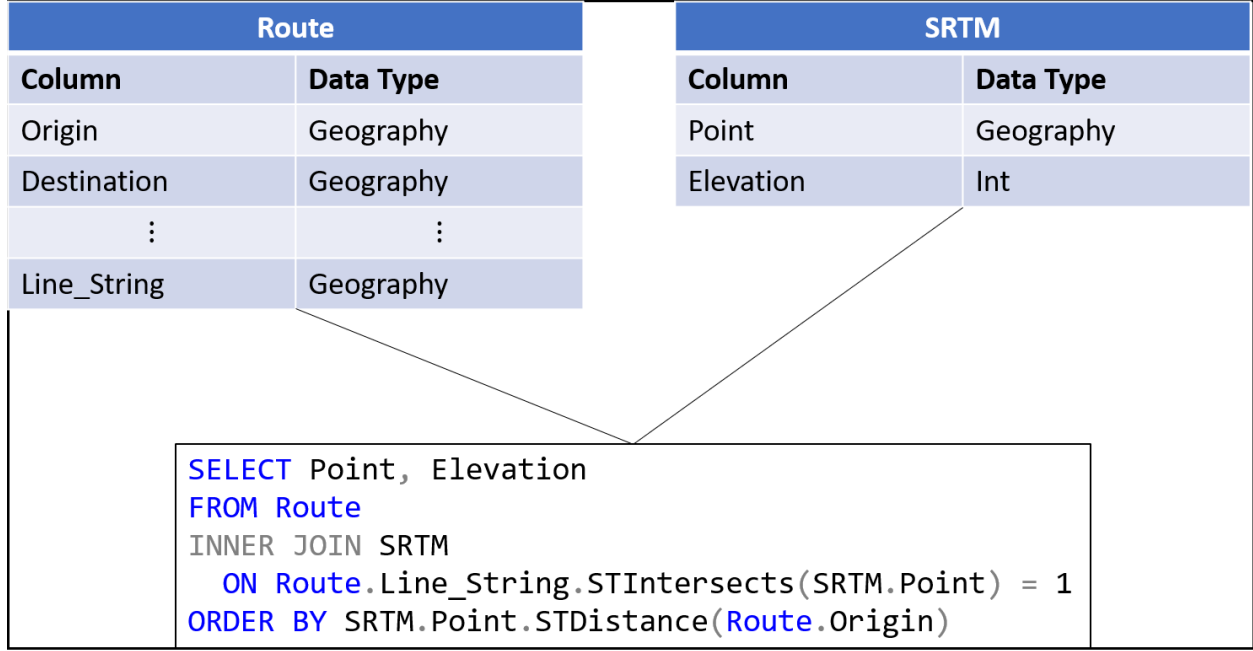


**Figure 14: Spatial view from SQL Server Management Studio showing line string object and intersecting SRTM coordinates.**

The SRTM data points are then joined to a route record using the STIntersects function, then ordered by distance from the route origin with STDistance function. These two functions leverage SQL Server's R-tree indexing, and average query time is under one second for a dataset with over 100 million SRTM records. Figure 15 shows an example query with representations of the data table structure.

---

<sup>1</sup>Microsoft official documentation of spatial functions can be found [here](#).



**Figure 15: Example query showing how routing and SRTM elevation are joined using SQL Server spatial functions.**

From the points obtained using the query in Figure 15, an elevation factor is calculated using each SRTM elevation point similar to equation (11). Because the SRTM points crisscross over the line string, a weighted distance is used to better scale the contribution of distance between them:

$$d'_{st,j,j+1} = \frac{r_{st}}{\sum_{j=0}^{n_{st}} d(s_{st,j}, s_{st,j+1})} d(s_{st,j}, s_{st,j+1}) \quad (12)$$

where  $n_{st}$  is the total number of shape points obtained from the query in Figure 15. Substituting into equation (11), a new elevation factor is calculated as

$$\lambda_{st} = \sum_{j=0}^{n_{st}-1} \frac{|z_{ij,j+1} - z_{ij,j}|}{d'_{st,j,j+1}} \quad (13)$$

The final two models are provided in Table 8. Model 5 uses the average travel speed for the urban area as used in Model 1 and Model 3, whereas Model 6 uses the average travel speed calculated for each origin-destination pair as in Model 2 and Model 4.

**Table 8: Models tested on SRTM spatial data**

Model 5	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}} + \xi \bar{v}^2 \right) + b_\mu r_{st} \mu \eta + b_\lambda \bar{\lambda}_{st} \mu \eta$
Model 6	$F_{st} \approx b_v r_{st} \left( \frac{\gamma}{\bar{v}_{st}} + \xi \bar{v}_{st}^2 \right) + b_\mu r_{st} \mu \eta + b_\lambda \bar{\lambda}_{st} \mu \eta$

### 4.3 Experimental Results

An OSM data extract for the state of Oregon was obtained from Geofabrik (Karch & Ramm, 2018) and processed with SRTM data obtained from the USGS website (USGS, 2018). The data for the pathfinding engine and subsequent analyses were stored in a SQL Server 2017 Enterprise<sup>2</sup> database installed on a virtualized Windows Server 2016 with 32GB of RAM and 8 virtual CPUs clocked at 3.47GHz. SQL Server's built-in spatial indexing was utilized for geometrically calculated independent variables and all spatial objects are stored as geography data types.

The urban test areas were chosen with 20 hand-picked locations producing 380 path calculations (i.e. all-pairs calculations from each location to all others). Additionally, vehicle mass was varied between 6,350 kg and 36,350 kg in increments of 5,000 kg. Table 9 provides summary statistics for the urban areas tested.

Table 10 provides summary statistics for the independent variables described in sections 4.2.2 and 4.2.3.

**Table 9: Urban area statistics**

Urban Area	Avg. Network Travel Speed $\bar{v}$ (km/hr.)	Avg. Elevation Change $\alpha$	Area (km <sup>2</sup> )	Min. Elevation (m)	Max. Elevation (m)	Orig/Dest pairs
Eugene	29.41	0.12	809.58	100	680	380
Portland	35.30	0.107	1579.47	-18	432	380
Salem	32.23	0.147	804.7	32	351	380

<sup>2</sup> It is notable that free Developer edition of SQL Server 2017 also provides the same spatial functionality.

**Table 10: Independent variable summary statistics**

Variable	Units	Min	Max	Mean	Median	Std. Dev
$\mu$	Kilograms	6350.00	36350.00	21350.00	21350.00	10000.63
$r_{st}$	Meters	2303.43	47202.70	14809.08	12761.78	8777.83
$v_{st}$	Meters/second	26.23	36.32	32.04	31.88	2.37
$\alpha_{st}$	Dimensionless	0.0890	0.1717	0.1238	0.1187	0.0194
$\lambda_{st}$	Dimensionless	0.0000	0.6073	0.0608	0.0462	0.0573

To evaluate predictive accuracy of the FCEFs, the mean percentage error (MPE) and mean absolute percentage error (MAPE) are calculated as

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{F_i - \hat{F}_i}{F_i} * 100\%,$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{F_i - \hat{F}_i}{F_i} \right| * 100\%$$

where  $F_i$  is the actual fuel consumption of the fuel-efficient route and  $\hat{F}_i$  is the predicted value. MPE indicates whether the model on average underestimates or overestimates the actual value; MAPE provides the average deviation as a percentage of the actual value and is overall indicative of the predictive accuracy of the model.

#### 4.3.1 Benchmarking

To provide a reference for benchmarking the MAPE values of the developed FCEFs, simple circuitry factors are calculated for the distance traveled along a shortest path, a fastest path and a fuel-efficient path for all origin-destination pairs. The regression formulas are the same as equation (2) in the literature review with the constant forced to zero:

$$D_{st} \approx \beta r_{st} \quad (14)$$

where  $D_{st}$  is the estimated travel distance between an origin  $v_s$  and destination  $v_t$ ;  $r_{st}$  is the Euclidean distance; and  $\beta$  is the parameter to estimate. Table 11 provides regression results for equation (14). It's notable that the *distance* of the fuel-efficient path can be estimated with similar accuracy as paths optimized for shortest distance and least travel time.

**Table 11: Regression results for DEFs**

Path Optimization Type	Aggregation	$\beta$	t-stat	$R^2$	MPE	MAPE
Shortest	Pooled	1.219	1025.57	0.9925	3.68%	8.10%
	Eugene	1.218	492.63	0.9892	4.95%	9.26%
	Portland	1.220	698.19	0.9946	1.68%	6.65%
	Salem	1.217	503.72	0.9896	4.57%	8.44%
Fastest	Pooled	1.345	705.72	0.9842	2.87%	10.80%
	Eugene	1.326	359.24	0.9798	5.04%	11.47%
	Portland	1.357	453.31	0.9872	1.67%	9.48%
	Salem	1.326	362.68	0.9802	3.65%	11.47%
Fuel-efficient	Pooled	1.327	696.45	0.9838	4.16%	10.74%
	Eugene	1.318	372.16	0.9812	5.50%	11.35%
	Portland	1.326	426.84	0.9856	2.74%	9.61%
	Salem	1.341	374.68	0.9814	4.05%	11.33%

#### 4.3.2 Graph dataset results

Table 12 summarizes results for each of the four models pooled as well as aggregated by each urban test area. It is encouraging that all estimated parameters are not only highly statistically significant but also make intuitive sense by being positively correlated with increasing distance, vehicle mass, and absolute elevation change. Although vehicle weight and straight-line distance appear to be reasonably good estimators of fuel consumption, the inclusion of variables for average absolute elevation change in models 3 and 4 ( $\bar{\alpha}$  and  $\bar{\alpha}_{st}$ , respectively) provides an improvement in the predictive accuracy as shown by the reduced MAPE. Furthermore, per-path queries of average travel speed ( $\bar{v}_{st}$ ) and elevation change ( $\bar{\alpha}_{st}$ ) provide improvement for models 2 and 4 over 1 and 3, respectively for the pooled dataset.

It is notable that results improve when the coefficients are estimated by urban area and the MAPE values approach those of the DEFs in section 4.3.1.

**Table 12: Regression results for FCEFs derived from the graph dataset**

Aggregation	Model	$b_v$	t-stat	$b_\mu$	t-stat	$b_\alpha$	t-stat	$R^2$	MPE	MAPE
Pooled	1	1.1553	39.45	0.0446	126.57	-	-	0.9478	11.74%	21.70%
	2	1.1756	39.96	0.0445	126.36	-	-	0.9480	11.71%	21.67%
	3	1.2380	48.70	0.0337	90.76	7.5143	51.49	0.9608	-1.60%	18.40%
	4	1.2577	49.17	0.0339	91.66	7.3805	51.17	0.9609	-1.38%	18.45%
Eugene	1	4.5616	48.78	0.0427	74.04	-	-	0.9679	3.81%	16.62%
	2	4.4883	50.82	0.0427	76.75	-	-	0.9691	4.36%	16.73%
	3	4.5616	51.26	0.0361	53.08	3.8995	16.69	0.9709	-2.62%	15.66%
	4	4.5129	54.20	0.0358	55.43	4.1474	18.26	0.9726	-2.45%	15.47%
Portland	1	1.3851	27.44	0.0422	94.12	-	-	0.9676	5.00%	15.56%
	2	1.3425	25.89	0.0428	94.51	-	-	0.9668	5.67%	15.84%
	3	1.3851	31.76	0.0309	57.36	11.9867	30.05	0.9758	-2.63%	14.25%
	4	1.3899	32.33	0.0308	60.42	12.3548	34.84	0.9772	-2.87%	14.04%
Salem	1	2.3090	26.87	0.0421	66.03	-	-	0.9455	9.48%	19.66%
	2	2.3141	26.73	0.0422	66.48	-	-	0.9454	9.79%	19.74%
	3	2.3090	33.24	0.0271	41.77	7.1504	37.57	0.9644	-4.37%	17.95%
	4	2.3609	33.55	0.0270	40.95	7.0557	36.99	0.9639	-4.11%	18.03%

### 4.3.3 SRTM dataset results

Results for the SRTM dataset models are presented in Table 13. These models provide improvements over Models 1 and 2, however the substitution of  $\bar{v}_{st}$  for  $\bar{v}$  to better estimate per-path travel offers no improvement. This suggests that further improvements in model accuracy are likely to be obtained through better estimations of the impact of elevation change.

The high statistical significance and intuitive values of the estimated parameters along with a relatively low MAPE demonstrate the utility of the DEM approach that may be more straightforward to calculate in a real-life planning context. The approach is also arguably more parsimonious by measuring properties along the Euclidean line connecting the origin and destination similar to linear DEFs.



**Table 13: Regression results from SRTM dataset**

Aggregation	Model	$b_v$	t-stat	$b_\mu$	t-stat	$b_\lambda$	t-stat	$R^2$	MPE	MAPE
Pooled	5	1.444	34.07	0.0443	120.45	4.063	23.77	0.9479	9.34%	20.99%
	6	1.504	35.12	0.0439	119.72	4.109	24.12	0.9483	9.25%	20.97%
Eugene	5	4.562	49.21	0.0410	65.90	1.797	6.97	0.9684	1.63%	16.31%
	6	4.502	51.60	0.0407	67.36	2.054	8.15	0.9699	1.84%	16.36%
Portland	5	1.385	30.49	0.0358	74.77	25.051	24.98	0.9738	0.56%	14.08%
	6	1.349	28.85	0.0362	74.72	25.208	24.79	0.9730	1.16%	14.30%
Salem	5	2.309	26.96	0.0410	60.89	0.839	4.47	0.9459	8.71%	19.35%
	6	2.327	26.99	0.0409	60.49	0.979	5.21	0.9459	8.84%	19.36%

## **5.0 CONCLUSIONS**

### **5.1 Challenges**

This research was challenging for several reasons. The use of a custom-developed pathfinding engine required a considerable amount of time to program and configure with the OSM and SRTM datasets. A reasonable effort was put forth to determine the feasibility of using existing software, either for the path calculations or processing the elevation model. However, the modeling of fuel consumption costs required some flexibility with the routing algorithm. Fuel-efficient routing software also tends to be commercial and closed source, limiting its research applications.

The development of the regression models was also challenging because of the lack of theoretical research available. Fuel consumption is highly dependent on several vehicle-dependent and environmental variables, so any research must carefully isolate a handful of these to develop meaningful relationships and results. In that regard, the vehicle attributes were chosen somewhat arbitrarily but are based on those of a typical over-the-road commercial vehicle with gross vehicle weight (GVW) of 80,000 lbs.

Working with raw SRTM data also presented some challenges over a more refined data source. The SRTM data are still relatively patchy in areas and are often augmented with other data sources to fill in gaps. While developing the experimental setting, some effort was put forth to identify data gaps and avoid them when choosing path origin-destination locations. However, large-scale refinement of the data was outside the scope of this research.

### **5.2 Recommendations and future work**

There are several ways this research could be improved and expanded. The first and likely most important would be validation of the fuel consumption costs calculated using real consumption data or drive cycle simulations. Several improvements to the pathfinding engine can also be made to account for idling, acceleration and turning movements, all of which increase consumed fuel. The pathfinding engine is capable of calculating turn costs in the context of a fastest path calculation, however extending this to fuel consumption was outside the scope of this research.

As mentioned prior, there is much room for improvement with the elevation dataset used. The SRTM data were chosen for their availability and straight-forward format. However, the USGS maintained DEM dataset is more accurate and available in more detailed 1/3 and 1/9 arc-second resolution. For elevation modeling of the lower 48 U.S. states this is often the first choice. Nonetheless, the data may require some processing to remove anomalies and inaccuracies that are inevitably present in raw DEM datasets.

Additional sensitivity analysis could be conducted to determine the response of the dependent variable to changes in vehicle attributes. For example, a commercial operator may be interested in determining the fuel cost savings of choosing a particular vehicle spec for a terminal location based on expected customer delivery locations. Such an analysis would best be conducted by first validating the models against a range of engine specifications (e.g. the displacement  $V$  and engine nominal RPM  $N_e$  from equation (7)) and perhaps different aerodynamic body designs (e.g. frontal area  $A$  and coefficient of drag  $C_d$ ) with actual or simulated consumption data. Although it is obvious that smaller engine displacement and more aerodynamic body designs will reduce fuel costs, the choice often involves significant trade-offs. A smaller engine spec may reduce a commercial operator's ability to haul certain loads in a given urban area and aerodynamic body designs generally increase the cost of the vehicle or add components (e.g. trailer skirts) that are more easily damaged in day-to-day operations, thus increasing operating costs. Overall, reframing the analysis to assess fleet cost savings for carriers is entirely feasible and a worthwhile extension of this research.

A final minor improvement would also be the establishment of benchmark distributions of locations to generate paths between. In this research locations were basically hand-picked while attempting to cover the urban area as evenly as possible. Establishing methodology for this would likely improve the quality of the FCEFs.

## 6.0 REFERENCES

- Abraham, I., Dellling, D., Goldberg, A. V. & Werneck, R. F., 2010. *A Hub-Based Labeling Algorithm for Shortest Paths on Road Networks*, Redmond, WA: Microsoft Research.
- Ballou, R. H., Rahardja, H. & Sakai, N., 2002. *Selected country circuitry factors for road travel distance estimation*. Washington, D.C., Transportation Research Board, p. 843–848.
- Barth, M., Younglove, T. & Scora, G., 2005. *Development of a Heavy-Duty Diesel Modal Emissions and Fuel Consumption Model*, Riverside: California PATH Research Report.
- Cappiello, A. et al., 2002. *A Statistical Model of Vehicle Emissions and Fuel Consumption*. Singapore, Intelligent Transportation Systems, pp. 801-809.
- Figliozzi, M. A., 2008. *Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer Demands and Routing Constraints*. Washington, D.C., Journal of the Transportation Research Board, pp. 1-8.
- Franceschetti, A., Honhon, D., Woensel, T. V. W. & Bektas, T., 2013. The time-dependent pollution-routing problem. *Transportation Research Part B*, pp. 267-269.
- Geisberger, R., Sanders, P., Schultes, D. & Dellling, D., 2008. Contraction hierarchies: Faster and simpler hierarchical routing in road networks. *Experimental Algorithms*, pp. 319-333.
- Goldberg, A. V. & Harrelson, C., 2005. *Computing the Shortest Path: A\* Search Meets Graph Theory*. Philadelphia, PA, Society for Industrial and Applied Mathematics, pp. 156-165.
- Gonçalvesa, D. N. S., Gonçalvesb, C. d. M., Assisc, T. F. d. & Silvad, M. A. d., 2014. *Analysis of the difference between the Euclidean distance and the actual road distance in Brazil*. Sevilla, Spain, s.n.
- Karch, C. & Ramm, F., 2018. *Geofabrik Downloads: OpenStreetMap Data Extracts*. [Online] Available at: <http://download.geofabrik.de/> [Accessed 4 February 2018].

Lopp, S., Wood, E. & Duran, a. A., 2015. Evaluating the Impact of Road Grade on Simulated Commercial Vehicle Fuel Economy Using Real-World Drive Cycles. *SAE Commercial Vehicle Engineering Conference*, pp. 3-5.

OpenStreetMap Wiki contributors, 2017. *OSM XML*. [Online] Available at: [https://wiki.openstreetmap.org/wiki/OSM\\_XML](https://wiki.openstreetmap.org/wiki/OSM_XML) [Accessed 27 April 2018].

OpenStreetMap Wiki contributors, 2018. *O5m*. [Online] Available at: <http://wiki.openstreetmap.org/w/index.php?title=O5m&oldid=1594164> [Accessed 16 May 2018].

OpenStreetMap Wiki contributors, 2018. *PBF Format*. [Online] Available at: [http://wiki.openstreetmap.org/w/index.php?title=PBF\\_Format&oldid=1580310](http://wiki.openstreetmap.org/w/index.php?title=PBF_Format&oldid=1580310) [Accessed 3 June 2018].

Pfoser, D., Efentakis, A., Voisard, A. & Wenk, C., 2009. Exploiting Road Network Properties in Efficient Shortest-Path Computation. *International Computer Science Institute*, pp. 4-7.

Schaper, T. & Bruns, C., 2015. *Fuel efficient route calculation*, s.l.: European Commission Seventh Framework Programme.

Shahid, R., Bertazzon, S., Knudtson, M. L. & Ghali, W. A., 2009. *Comparison of distance measures in spatial analytical modeling for health service planning*, s.l.: BMC Health Services Research.

USGS, 2018. *Index of SRTM3*. [Online] Available at: [https://dds.cr.usgs.gov/srtm/version2\\_1/SRTM3/North\\_America/](https://dds.cr.usgs.gov/srtm/version2_1/SRTM3/North_America/) [Accessed 12 March 2018].

Wood, E., Burton, E., Duran, A. & Gonder, J., 2014. *Appending High-Resolution Elevation Data to GPS Speed Traces for Vehicle Energy Modeling and Simulation*, Denver, CO: National Renewable Energy Laboratory.

Wood, E., Burton, E., Duran, A. & Gonder, J., 2014. Contribution of road grade to the energy use of modern automobiles across large datasets of real-world drive cycles. *SAE World Congress*, pp. 2-4.

Wyatt, D. W., Li, H. & Tate, J. E., 2014. *The impact of road grade on carbon dioxide (CO<sub>2</sub>) emission of a passenger vehicle in real-world driving*. s.l., Transportation Research Board, pp. 160-170.

## 7.0 APPENDIX

### 7.1 Glossary of Terms

Term	Notation	Description
DEF	N/A	Distance-estimating function; a formula estimated using linear regression for calculating distances between origin-destination pairs in lieu of pathfinding software.
DEM	N/A	Digital Elevation Model; a 3-dimensional dataset consisting of geographic coordinates and associated elevation values (usually in reference to sea level).
Edge	$e_{ij}$	Represents connections between vertices in a graph as well as the relative cost of traversing from one vertex to another reachable vertex. An edge $e_{ij}$ connects vertices $v_i$ and $v_j$ .
FCEF	N/A	Fuel-consumption estimating function, a proposed extension of the distance-estimating function.
Shape point	$p_{ij,k}$	A non-decision point geographic coordinate along an edge used for storing elevation data. A given shape point has an elevation value $z_{ij,k}$ . In mapping applications shape points are used for rendering edges and storing alternate data.
SRTM	N/A	Shuttle Radar Topography Mission; an international research effort that generated digital elevation models for most of the world using specialized radar imaging equipment flown on the space shuttle Endeavor.
Vertex	$v_i, v_j$	A coordinate representing a decision point in a graph structure.

## 7.2 Urban Test Areas

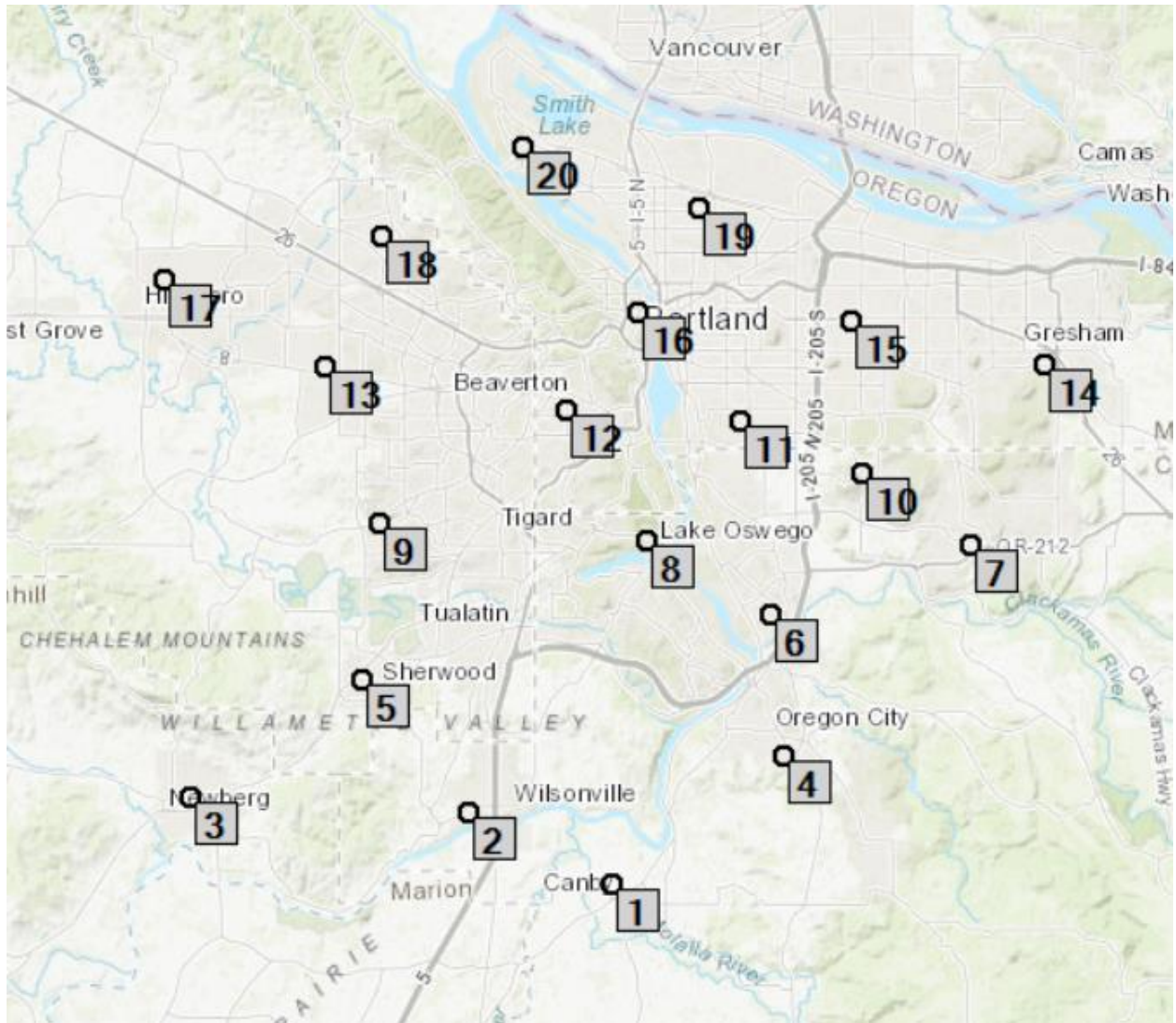
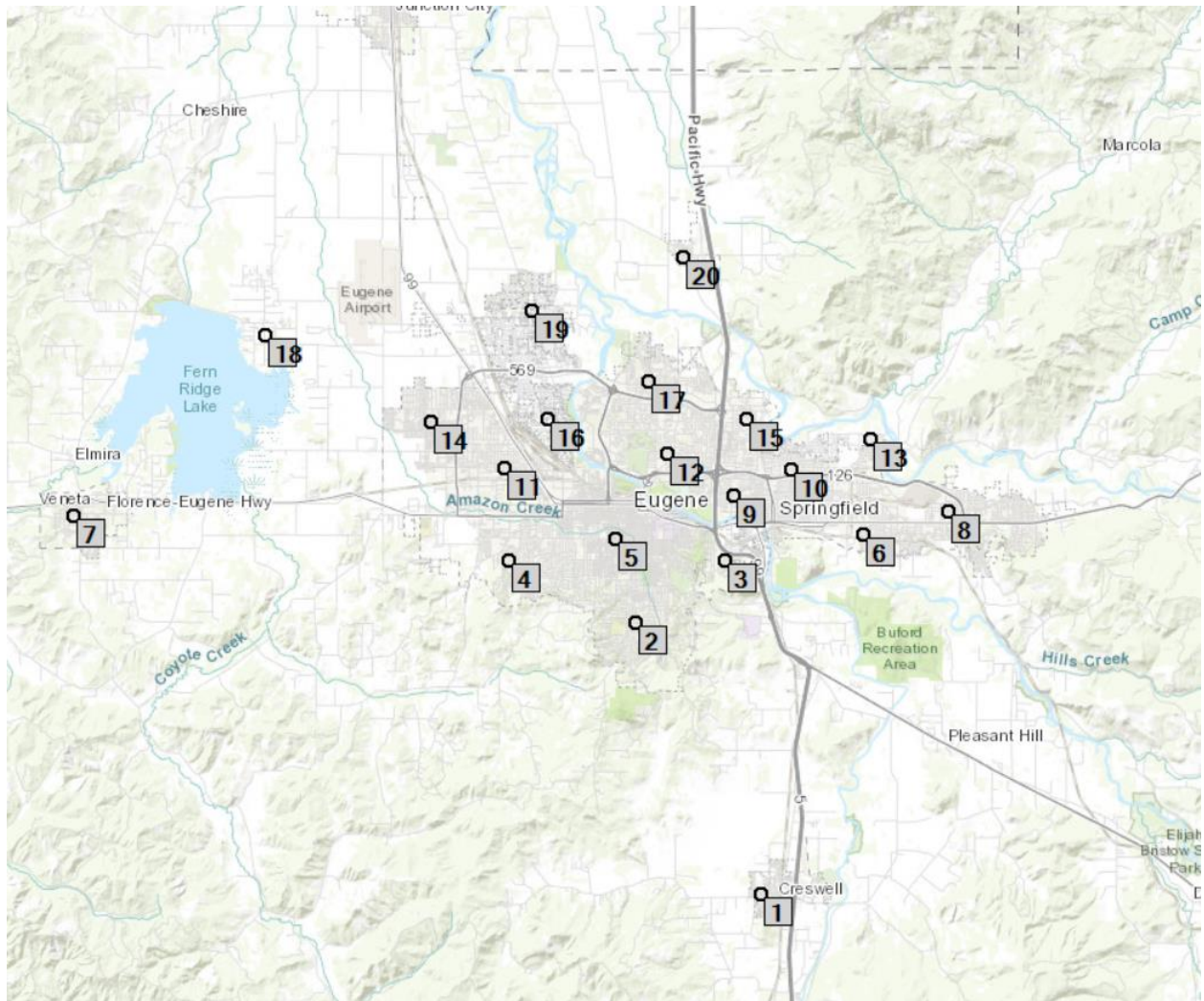
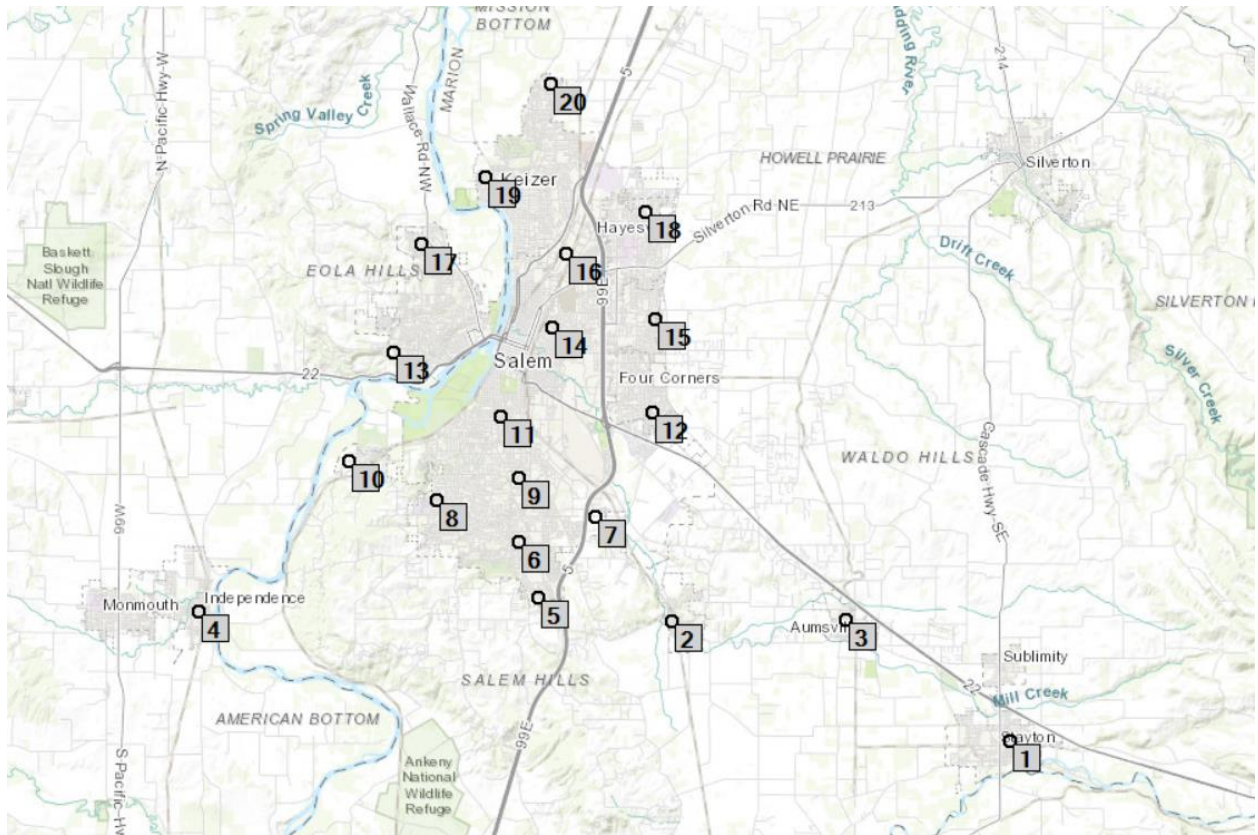


Figure 16: Portland, OR test area with path origin-destination locations.





**Figure 17: Eugene, OR test area.**



**Figure 18: Salem, OR test area.**